Article

Not peer-reviewed version

The Chloroplast Genome of a V. amygdalina Specie Isolated from Awka, Nigeria: Comparative and Phylogenetic Analysis.

Ogechukwu Nworji * , Joy Ogana , Ejike Orji , ikenna ohanenye , Obiajulu Ezeigwe , Azubuike Enemor

Posted Date: 12 July 2023

doi: 10.20944/preprints202307.0778.v1

Keywords: V. Amygdalina; Chloroplast genome; Medicinal Plant; Geographical location; Gene polymorphism.

Article

The Chloroplast genome of a V. amygdalina specie isolated from Awka, Nigeria: Comparative and Phylogenetic analysis.

Ogechukwu Nworji [1,*], Joy Ogana [1,*], Ejike Orji [1], Ikenna Ohanenye [2] and Obiajulu Ezeigwe [1] and Azubuike Enemor

[1] Department of Applied Biochemistry, Nnamdi Azikiwe University
[2] School of Nutrition Sciences, Florida State University (London Study Centre)
* Correspondence: of.nworji@unizik.edu.ng. Tel: +2348145671971

Abstract: *Vernonia amygdalina* is an important nutritional and medicinal plant that is used to treat various diseases such as malaria and diabetes in Nigeria. In this study, we sequenced, assembled, and annotated the complete nucleotide sequence of the chloroplast genome of a V. amygdalina species isolated in Awka, Nigeria. The chloroplast genome is 153 149 bp in length and has a conserved structure that includes the LSC (84261 bp) region, the SSC (13152 bp) region, and the IR (27868 bp) regions. A total of 131 genes were identified, including 86 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Comparative analysis of the genome revealed species variations in the LSC region associated with the geographical location of the V. amygdalina plant. Phylogenetic analysis of the chloroplast genomes using FastTree showed that *V. amygdalina* clustered closest to the chloroplast of the *Crepidiastrum* genus. This study identifies the characteristics of the *V. amygdalina* chloroplast genome, which will improve our understanding of intraspecies diversity and the identification of potential molecular markers to determine geographical origin.

Keywords: *V. Amygdalina*; Chloroplast genome; Medicinal Plant; Geographical location; Gene polymorphism

Introduction

Vernonia amygdalina is a tropical shrub of the *Astaraceae* family that is native to West Africa. It is commonly called bitter leaf because its leaves have a characteristically bitter taste, which can be reduced by boiling or wringing in cold water. *V. amygdalina* grows widely across West Africa, and in Southeastern Nigeria, it is commonly farmed on homesteads. The plant is economically valuable for its nutritional and medicinal properties. Nutritionally, *V. amygdalina* leaves are mainly used as a culinary herb staple in soups and stews across West Africa [1]. Ethnomedicinally, it has demonstrated antimicrobial [2, 3, 4], antidiabetic [5, 6], antiproliferative (4, 6), and anti-inflammatory activities [4]. Its efficacy against diverse tropical diseases is due to the presence of bioactive secondary metabolites in various parts of the plant. Beneficially, no toxic effect has been associated with its use as a therapeutic agent or adjuvant, even at above-optimal concentrations in animal models [3]. Intriguingly, the geographical location of *V. amygdalina* plants affects their constitutive bioactive compounds [1]. Understanding this phenomenon requires knowledge of *V. amygdalina* genetics. This study assembled the complete chloroplast genome of a *V. amygdalina* species grown in Awka, Nigeria.

The chloroplast is a photosynthetic organelle that provides essential energy and nutrients for photoautotrophic plants. This essential organelle is core to most metabolic processes in plant cells, including the synthesis of metabolites, stress responses, and growth regulation. The chloroplast (cp) genome encodes the genetic information required for photosynthesis and other biochemical processes in plant cells [7]. Characterising the CP genome is important for improved understanding of plant biology, phylogeny, and chloroplast genomic engineering [7]. The chloroplast genome is highly conserved compared to the plant nuclear and mitochondrial genomes. It has a low rate of recombination and is inherited maternally in most angiosperms, making it a great source of

 doi:10.20944/preprints202307.0778.v1

information for phylogenetic and population genetic analyses [8]. However, chloroplast genomic sequences show significant sequence variation within plant species from different geographical locations. The CP genome diversity within species provides important information on selective breeding to conserve valuable traits, plant adaptation to climate change, and the development of efficient species-specific chloroplast vector systems for genetic engineering [9, 10]. The cp genome in angiosperms is mostly circular double-stranded molecules with a highly conserved quadripartite structure consisting of a large single copy (LSC) region, a small single copy (SSC) region, and two copies of an inverted repeat (IR) region [9, 10].

In this study, we describe details of the de-novo assembly, annotation, and comparative analysis of the chloroplast genome of a *V. amygdalina* Delile specie grown in Awka, Nigeria. The results will improve understanding of the CP genomes of *V. amygdalina* and the *Asteraceae* family.

Results

Chloroplast Genome Structure of Vernonia Amygdalina

The complete chloroplast genome of *V. Amygdalina* (Nigeria) is 153149 bp. It displays the conserved quadripartite structure of chloroplast genomes, featuring an LSC region of 84261 bp, two copies of the IR region (27868 bp), and the SSC region of 13152 bp (Table 1). The GC content of the LSC region (35.8%), IR region (41.8%), and SSC region (31.8%) shows a pattern present in most chloroplast genomes, with the IR region having the highest GC content. The *V. Amygdalina* Cp genome consists of 131 genes, which include 86 genes, 37 transfer RNA (tRNA), and 8 ribosomal (rRNA) genes.

Table 1. *V. amygdalina* Chloroplast genome feature.

Region	*V. Amygdalina* (Nigeria)		Reference (NC_035143)	
	Size (bp)	%GC content	Size (bp)	%GC content
Cp Genome	153149	37.7	153133	37.7
LSC	84261	35.8	84245	35.8
SSC	13152	31.8	13152	31.8
IR	27868	41.8	27868	41.8

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2023 doi:10.20944/preprints202307.0778.v1

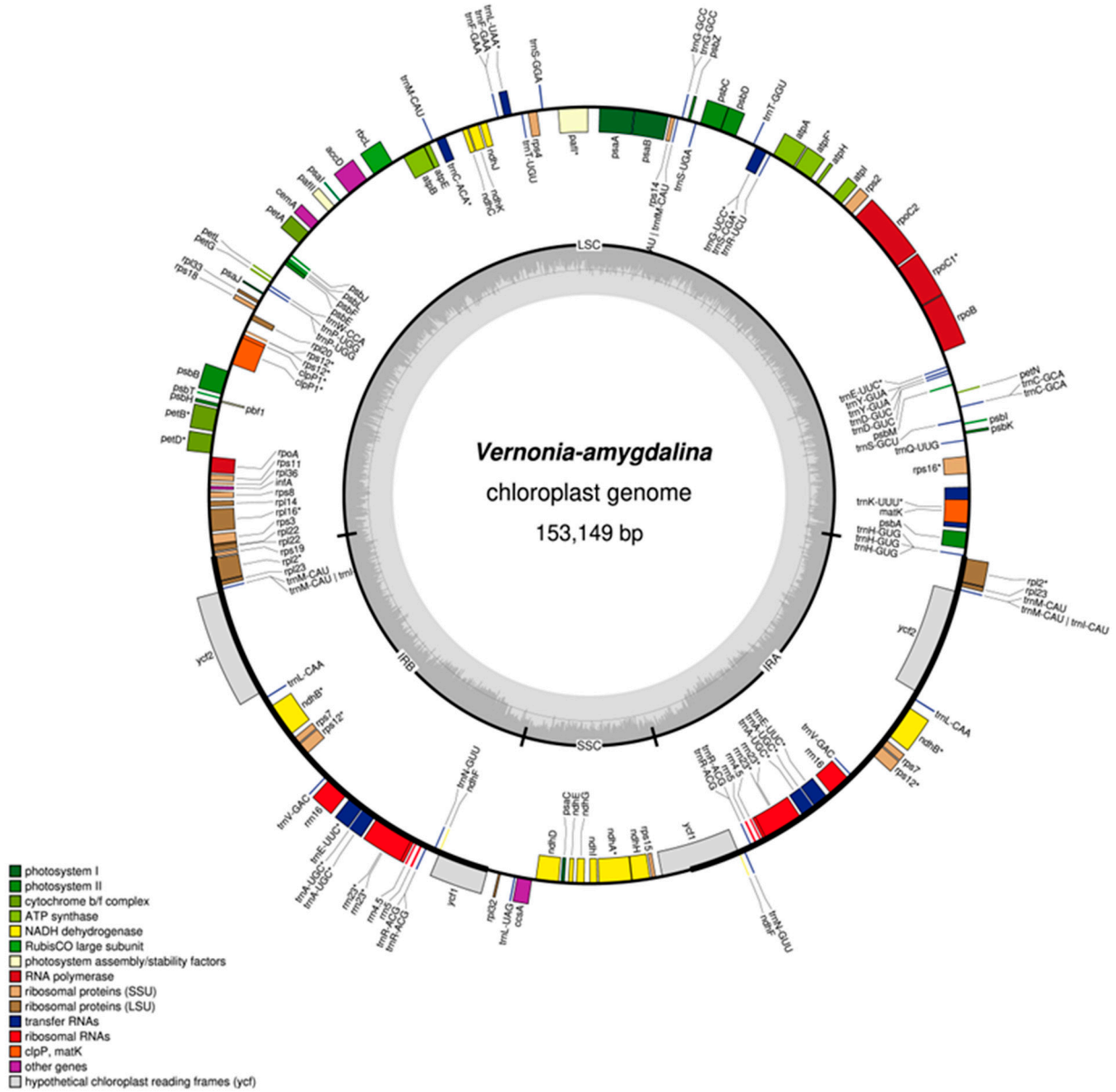

Figure 1. Map of *V. amygdalina* (Nigeria) chloroplast genome Genes shown outside the map are transcribed clockwise, while the genes shown inside are transcribed anticlockwise. The dark grey area of the inner circle indicates GC content, while the lighter area indicates AC content. Genes of different functional groups are colour coded. LSC, large single copy region; IR, inverted repeat; SSC, small single copy region.

Comparative Analysis of the V. amygdalina Chloroplast Genome

SNP and Indel analyses were performed on the cpDNA sequence of *V. amygdalina* using the reference sequence NC_053851 to detect intraspecies gene diversity in the cp genomes. The sequence variation mainly occurred in the intergenic regions of the LSC and SSC loci of the chloroplast genome. However, two mutations affected the genes: the ndhA exon 2 in the SSC region was mutated, resulting in an amino acid change (isoleucine to valine), while a silent mutation affected the psbA gene in the LSC region (Table 2). A total of fourteen mutations (6 SNPs and 8 indels) were detected; twelve mutations occurred in the LSC region and one in the SSC region.

Table 2. Variants and SNPs in the *V. amygdalina* Chloroplast genome.

Name	Change	Codon Change	A. Change	A. Polymorphism Type	Gene	Protein Effect

ATAAA	+ATAAA			Insertion		
G	G -> A	TTT -> TTC		SNP (transition)	psbA	None
A	(A)9 -> (A)10			Insertion (tr)		
A	T -> A			SNP (transversion)		
TATAGA	+TATAGA			Insertion		
A	G -> A			SNP (transition)		
	(T)12 -> (T)11			Deletion (tr)		
TATAAG	+TATAAG			Insertion		
G	T -> G			SNP (transversion)		
C	A -> C			SNP (transversion)		
	(T)15 -> (T)14			Deletion (tr)		
T	(T)11 -> (T)12			Insertion (tr)		
	(T)14 -> (T)13			Deletion (tr)		
C	T -> C	ATT -> GTT	I -> V	SNP (transition)	ndhA	Substitution

* **A.A.** Amino acid; **tr** tandem repeats.

Phylogenetic Analysis

In this study, a phylogenetic tree was constructed with the complete chloroplast sequence of *V. amygdalina*, the reference sequence, and twenty unique species from the *Asteraceae* to determine the evolutionary relationship within the family. The analysis indicates a close genetic relationship between the *Vernonia* and *Crepidiastrum* genera. Interestingly, the *V. amygdalina* chloroplast genome shared a common monophyletic node with the reference sequence, even with fourteen variations to the reference site.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 July 2023 doi:10.20944/preprints202307.0778.v1

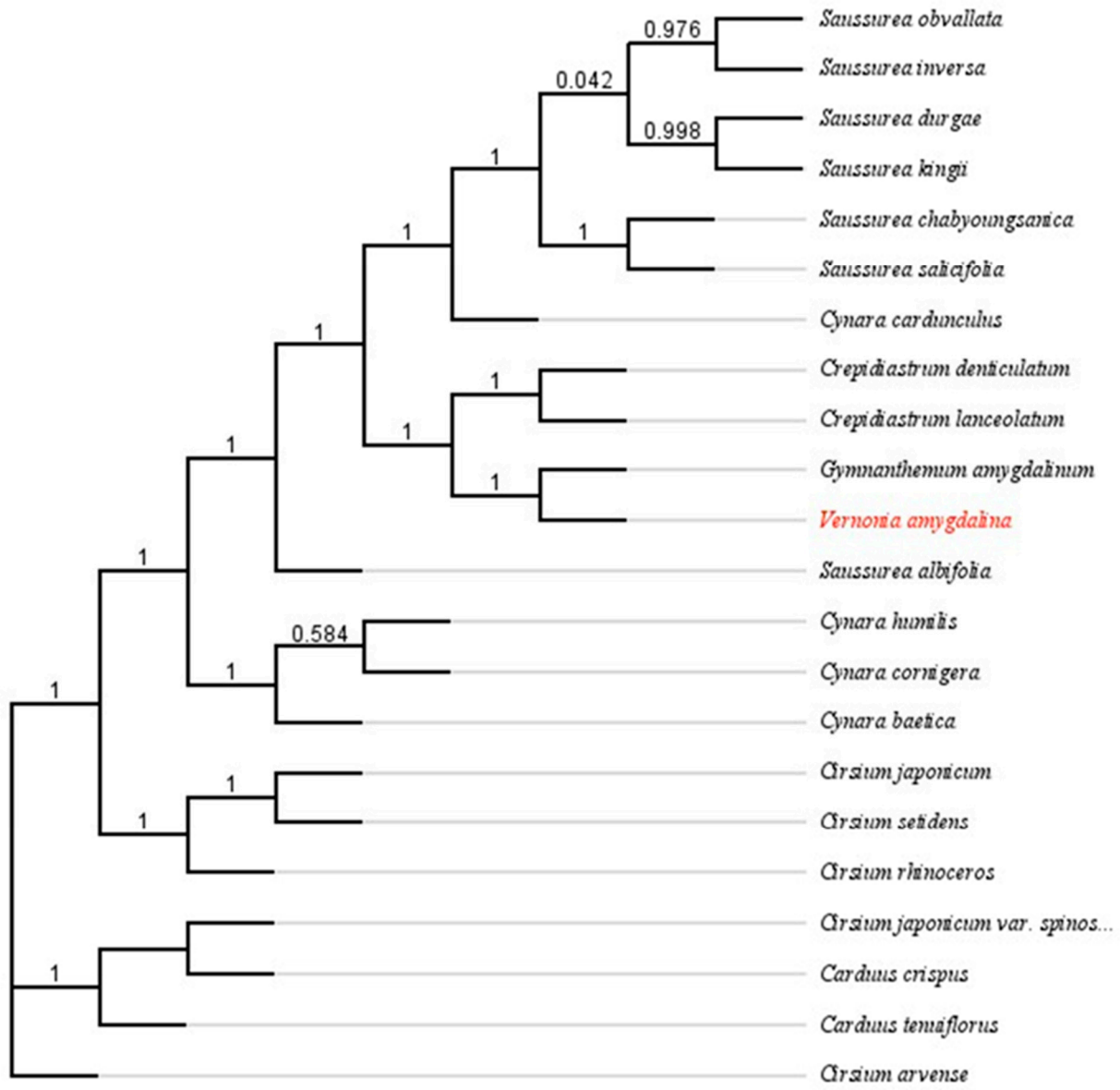

Figure 2. Phylogenetic Tree constructed using FastTree based on the chloroplast genomes of 21 different species The numbers represent the Fast Tree support value.

Discussion

The chloroplast genome of *V. amygdalina* showed a conserved quadripartite structure consisting of a large single copy (LSC) region, a small single copy (SSC) region, and two copies of an inverted repeat (IR) region [9, 10]. The *V. amygdalina* genomes have identical GC content, although there was an expansion in the LSC region of the Nigerian variety. The extension of the LSC intergenic region introduced two new six-nucleotide motifs (TATAGA and TATAAG) that resulted in tandem repeats on the genome. Changes to the intergenic space of the LSC region are a common observation in chloroplast genomes [9]. In addition, the presence of SNPs and indels indicates that there could be considerable intraspecific variation present in the chloroplast genomes of *V. amygdalina* varieties and that the LSC region is potentially a good source of molecular markers to detect intraspecific cpDNA sequence diversity associated with geographical origin. No variation occurred in the IR loci, indicating that the IR region of V. amygdalina is conserved. This conservation of the IR genome within species has been observed in pangenomic studies of cucumber [11], yam [12], and tobacco [13] chloroplast genomes.

The ndh genes were conserved during the evolution of cyanobacteria from an endosymbiont to present-day chloroplasts in plants. The chloroplast of most photosynthetic plants contains eleven ndh genes, which encode components of the thylakoid ndh complex that modulate the redox level of the cyclic photosynthetic electron transporters (14). This modulation is necessary as damage by ROS produced by excess light would impair photosynthesis. In other words, the ndh complex protects against photooxidative-related stress while maintaining an optimal rate of cyclic photophosphorylation (15).

Post-transcriptional editing of cytosine to uracil is a feature of plastids that is essential for correct gene expression. In plastids, RNA editing mostly reverts the transcripts back to the coding of conserved amino acids, resulting in a functional protein that has a different amino acid sequence from its analogous DNA sequence (15, 16). Several editing sites are located on the NDH genes. The prevailing hypothesis for this phenomenon is that the ndh genes in angiosperm ancestors were in the nucleus after endosymbiosis, where they accumulated a lot of T-to-C inactivating mutations. During evolution, perhaps in a high abiotic stress condition, they were relocated back to the chloroplast genome, where the C bases are corrected back to U in transcripts and with time in the DNA (15, 16). Therefore, the T-to-C mutation at base 264 of ndhA-2 may not have any profound effect on the plant. However, this requires further investigation.

Chloroplast genomes are widely used for phylogenetic analysis because they are mostly inherited maternally and are highly conserved compared to plant nuclear and mitochondrial genomes [9, 10]. Therefore, it is suitable for mapping out the diversity and plasticity of plants. In other words, when there is a significant variation within a species, phylogenetics can be used to determine the geographical origins of variants. Understanding intraspecific variation is essential, especially for medicinal plants, as intraspecific variation in secondary metabolites is influenced by geographical terrain [17, 18].

In this study, we assembled the chloroplast genome of V. amygdalina from Awka, Anambra State, and provided insight into the phylogeny of the *Asteraceae* family. The results provide important information on the intraspecific genetic variation present in V. amygdalina. The LSC region is remarkably variable in V. amygdalina and therefore can be a potential source for molecular markers for geographical species identification. The data provides a molecular basis for further exploration of genetic variation in the V. amygdalina population and its impact on breeding and metabolism.

Materials and Methods

Plant Material, DNA Extraction, and Sequencing

Fresh *V. amygdalina* leaves were obtained from Nnamdi Azikiwe University. Intact chloroplasts were extracted using the Minute Chloroplast Isolation Kit (Invent Biotechnologies, USA) according to the manufacturer's protocol. DNA was isolated from the chloroplast using the ZYMO Quick-DNA Plant/Seed Miniprep Kit according to the manufacturer's instructions. The isolated cpDNA was sent to the Inqaba Biotech laboratory in Ibadan for Next Generation Sequencing.

De Novo Assembly and Annotation

Sequence reads were assembled using the Geneious Prime software (Version 2023.1.2: https://www.geneious.com). The reads were paired and then trimmed using BBDuk to remove adapter sequences and low-quality reads with a quality value less than Q15. Trimmed reads were de novo assembled using the Geneious assembler. The contigs obtained were mapped to a reference sequence of *V. amygdalina* (NC_053851) [19]. The contigs were concatenated to complete the assembly. The assembled genome was annotated using GeSeq [20], and a circular map of the genome was produced using OGDRAW [21]. The software tools can be assessed at Chlorobox: https://chlorobox.mpimp-golm.mpg.de/index.html. The chloroplast genome of V. amygdalina was submitted to the NCBI database (GenBank accession number: OR197626; BioProject accession number: PRJNA989263).

Comparative Analysis of Chloroplast Genomes

The complete chloroplast genome of V. amygdalina isolated from Awka, Nigeria, was compared with the reference sequence (NC_053851). Single nucleotide polymorphisms (SNP) and indels were detected with Geneious Prime.

Phylogenetic Analysis

Twenty chloroplast genome sequences from *Asteraceae* were used for phylogenetic analysis to determine the phylogenetic position of *V. amygdalina*. Genome alignment was performed using MAFFT (version 7.490) [22], and a phylogenetic tree was produced using FastTree (version 2.1.11) [23] on Geneious Prime.

Author Contributions: *O.N. and *J.O. These two authors contributed equally to this work". Conceptualization, O.N., J.O., and E.O.; methodology, O.N., J.O., E.O., and A.E.; Formal Analysis O.N., J.O., O.E., and I.O.; resources, O.N., J.O., E.O., I.O., O.E., and A.E.; writing—original draft preparation, O.N., J.O., and A.E.; writing—review and editing, O.N., J.O., E.O., I. O., O.E., and A.E.; Supervision, O.N., and J.O.; project administration, O.N., J.O., and E.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The chloroplast genome of V. amygdalina was submitted to the NCBI database (GenBank accession number: OR197626; BioProject accession number: PRJNA989263).

Acknowledgments: We thank the Molecular Research Foundation for Student and Scientist (MRFSS) Laboratory, which kindly availed us of the use of their facility.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yeap, S. K., Ho, W. Y., Beh, B. K., Liang, W. S., Ky, H., Yousr, A. H. N., & Alitheen, N. B. Vernonia amygdalina, an ethnoveterinary and ethnomedical used green vegetable with multiple bioactivities. Journal of Medicinal Plants Research 4, 2787–2812 (2010).
2. Evbuomwan, L., Chukwuka, E. P., Obazenu, E. I. & Ilevbare, L. Antibacterial activity of Vernonia amygdalina leaf extracts against multidrug resistant bacterial isolates. Journal of Applied Sciences and Environmental Management 22, 17 (2018).
3. Oyeyemi, I. T., Akinlabi, A. A., Adewumi, A., Aleshinloye, A. O. & Oyeyemi, O. T. Vernonia amygdalina: A folkloric herb with anthelminthic properties. Beni-Suef University Journal of Basic and Applied Sciences 7, 43–49 (2018).
4. Tinrat, S. & Singhapol, C. Evaluation of Antioxidant and Antibacterial Activities of Vernonia Amygdalina Leaf Extracts as An Auxiliary in Natural Hair Shampoo. Asian J Pharm Clin Res 50–57 (2020) doi:10.22159/ajpcr. 2020.v13i10.38581.
5. Ogbuagu, E. O., Airaodion, A. I., Ogbuagu, U. & Airaodion, E. O. Effect of Methanolic Extract of Vernonia amygdalina Leaves on Glycemic and Lipidaemic Indexes of Wistar Rats. AJRIMPS 1–14 (2019) doi:10.9734/ajrimps/2019/v7i330122.
6. Djeujo, F. M., Stablum, V., Pangrazzi, E., Ragazzi, E. & Froldi, G. Luteolin and Vernodalol as Bioactive Compounds of Leaf and Root Vernonia amygdalina Extracts: Effects on α-Glucosidase, Glycation, ROS, Cell Viability, and In Silico ADMET Parameters. Pharmaceutics 15, 1541 (2023).
7. Liu, X.-F., Zhu, G.-F., Li, D.-M. & Wang, X.-J. Complete chloroplast genome sequence and phylogenetic analysis of Spathiphyllum 'Parrish'. PLoS ONE 14, e0224038 (2019).
8. Asaf, S. et al. Comparative analysis of complete plastid genomes from wild soybean (Glycine soja) and nine other Glycine species. PLoS ONE 12, e0182281 (2017).
9. Green giant—a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny - Daniell - 2021 - Plant Biotechnology Journal - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/pbi.13556.
10. Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E. & Burleigh, J. G. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evolutionary Biology 14, 23 (2014).

11. Xia L, Wang H, Zhao X, et al. Chloroplast Pan-Genomes and Comparative Transcriptomics Reveal Genetic Variation and Temperature Adaptation in the Cucumber. Int J Mol Sci. 2023;24(10):8943. Published 2023 May 18. doi:10.3390/ijms24108943
12. Cao J., Jiang D., Zhao Z., Yuan S., Zhang Y., Zhang T., Zhong W., Yuan Q., Huang L. Development of chloroplast genomic resources in Chinese Yam (Dioscorea Polystachya) Biomed Res. Int. 2018;2018:6293847. doi: 10.1155/2018/6293847
13. Wang S., Gao J., Chao H., Li Z., Pu W., Wang Y., Chen M. Comparative chloroplast genomes of Nicotiana Species (Solanaceae): Insights into the genetic variation, phylogenetic relationship, and polyploid speciation. Front. Plant Sci. 2022; 13:1–15. doi: 10.3389/fpls.2022.899252
14. Ma M, Liu Y, Bai C, Yong JWH. The Significance of Chloroplast NAD(P)H Dehydrogenase Complex and Its Dependent Cyclic Electron Transport in Photosynthesis. *Front Plant Sci*. 2021;12:661863. doi:3389/fpls.2021.661863
15. Martín M, Sabater B. Plastid ndh genes in plant evolution. *Plant Physiology and Biochemistry*. 2010;48(8):636-645. doi:1016/j.plaphy.2010.04.009
16. Mohammed T, Firoz A, Ramadan AM. RNA Editing in Chloroplast: Advancements and Opportunities. *Curr Issues Mol Biol*. 2022;44(11):5593-5604. doi:3390/cimb44110379
17. Yanqun Li, Dexin Kong, Ying Fu, Michael R. Sussman, Hong Wu. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. Plant Physiology and Biochemistry, Volume 148, 2020, Pages 80-89, ISSN 0981-9428, https://doi.org/10.1016/j.plaphy.2020.01.006.
18. San Jose-Maldia, L., Matsumoto, A., Ueno, S. et al. Geographic patterns of genetic variation in nuclear and chloroplast genomes of two related oaks (Quercus aliena and Q. serrata) in Japan: implications for seed and seedling transfer. Tree Genetics & Genomes 13, 121 (2017). https://doi.org/10.1007/s11295-017-1202-4
19. Zhou F, Lan K, Li X, Mei Y, Cai S, Wang J. The complete chloroplast genome sequence of Vernonia amygdalina Delile. Mitochondrial DNA B Resour. 2021;6(3):1134-1135. Published 2021 Mar 19. doi:10.1080/23802359.2021.1902411
20. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R and Greiner S (2017) GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45: W6-W11
21. Greiner S, Lehwark P and Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47: W59-W64
22. Price, M.N., Dehal, P.S. and Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one, 5(3), p.e9490.
23. Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), pp.772-780.